

NSA review  
completed

*Automation*

*DDI Report*

*Policy  
USAF Final*

*Rpt. on  
Mach. Trans.*

*automation*

Acting Director for  
Intelligence  
Room 7E44 Headquarters

Chief, DDI Management Staff  
Room 2F28 Headquarters

USAF review(s)  
completed.

Here is the final report of the  
Department of the Air Force Working Group  
on Machine Translation on which [redacted]  
and [redacted] of FBIS represented CIA.  
[redacted] was the Intelligence  
Community Staff representative.

STAT

You might care to glance at the  
multiple mentions marked in red of FBIS  
and the recent FBIS-sponsored Machine  
Translation Seminar.

The Working Group recommends approx-  
imately [redacted] in R&D funding for  
machine translation over the next 6 years.  
Other conclusions and recommendations can  
be found on pp 9-10.

Director, Foreign Broadcast  
Information Service

DIA review(s)  
completed.

MACHINE TRANSLATION

PROBLEM: Assess the effectiveness and utility of machine translation of textual materials for the intelligence process.

EXPLANATION OF TERMS:

- Human Translation (HT): The process by which a human transfers the meaning from the linguistic pattern of a source language to the linguistic pattern of an object language.
- Machine-Aided Translation (MAT): A computer-based system for the storage and retrieval of lexical information, designed to assist and improve the proficiency of the human translator.
- Machine Translation (MT): A computer process which accomplishes the translation function without human assistance except, as required for quality improvement, in a post-edit function. Current MT systems have lexical, syntactical, and semantic components which produce translations of a given level of quality. Based on the amount of human post-editing required to achieve a level of quality acceptable for publication as a finished product, MT systems can be characterized as second or third generation. Second generation systems require 25-30 percent post-editing whereas a third generation system would require 5-10 percent post-editing. Increased sophistication of lexical, syntactical, and semantic components and the addition of techniques to accommodate contextual and pragmatic information are required in a third generation system to resolve problems of style and ambiguity in journalistic/literary prose. Finally, the automation of input to the MT process must be included in the context of an MT system.
- Scientific and Technical (S&T) Prose: A style of writing largely constrained and regularized by a need to present a logically coherent discourse. It is, therefore, amenable to the logical formulations required by second generation MT systems.
- Journalistic/Literary Prose: A style of writing that lacks precision and conciseness and is largely unconstrained and informal. It is characterized by colloquialisms, idioms, proverbial expressions, metaphor, and other literary devices. Syntactic complexity is exemplified by structural inversions, ellipsis, and grammatically incomplete utterances. A third generation MT system must be able to solve fundamental linguistic problems which obscure meaning and comprehension to provide a quality translation of journalistic/literary materials.

---

OPR: AF/INYXP (Maj Baldauf/76440)  
7 May 1976

FACTS:

- Exploitation of foreign language materials can provide valuable intelligence regarding capabilities and intentions.
- Current MT systems are useful and cost/effective for translation of S&T materials.
- Current MT systems are based on a conceptual base that is now 20 years old.
- Current input techniques are a major constraint to cost/effectiveness and timely MT production.
- Principal emphasis is on the translation of Russian into English.
- A substantial volume of Russian material of potential intelligence value remains untranslated.
- The quality of human translation varies considerably and human translator proficiency must be improved.
- Manpower and fiscal constraints preclude a substantial increase in human translator resources in the Department of Defense.
- Automated techniques can improve the quality and efficiency of the human translation process.
- Indicative translations and keyword applications have a utility in the intelligence process.
- Current state-of-the-art of MT cannot produce high quality journalistic/ literary output without significant post-editing.

ASSUMPTIONS:

- There has been considerable development in computer technology and linguistics such that a new synthesis is possible which might produce improved translations on a more cost/effective basis.
- Continued enhancement of currently operational MT systems will reach a point of diminishing returns; the third generation MT system requires an advanced technological approach.

- Machine-aided translation techniques and advanced Optical Character Reader (OCR) technology afford the greatest potential for improving the quality and efficiency of the translation process in the near term.
- Development of a third generation MT system is a long-term evolutionary process best pursued by support of promising technological approaches.

CRITERIA:

- To have utility to the intelligence process, an MT system must:
  - Provide more timely translations.
  - Offset a lack of human translation expertise.
  - Be competitive with human translation.
  - Be responsive to the needs of the consumer with regard to completeness and quality.

BACKGROUND:

1. Efforts to apply automated data processing (ADP) techniques to the problem of language translation have been under way for over 20 years. The Air Force Systems Command's Foreign Technology Division (FTD) has developed a large-scale MT system for the translation of Russian language scientific and technical literature which satisfies the needs of FTD's analysts. Other initiatives have been undertaken in the Intelligence Community which address dictionary development and keyword selection. All such efforts are aimed at improving the capability to exploit foreign language materials for intelligence purposes and provide a more responsive method for exchange of data with other nations.
2. The requirement to translate an increasing amount of contemporary Soviet military and socio-political materials has generated renewed interest in MT. As a result of initiatives by the Deputy Assistant Secretary of Defense (Resources and Management) and by the Assistant Chief of Staff, Intelligence, USAF, the Defense Intelligence Agency (DIA) conducted a preliminary survey of community translation requirements and needs for MAT and MT. This survey was provided to the Air Force by DIA letter, subject: "Survey of Machine Translation Requirements," 23 Jan 1976. In addition, the Foreign Broadcast Information Service sponsored a seminar on machine translation, 8-9 Mar 1976. This seminar was attended by a wide variety of experts from government, industry, and the academic community and provided a valuable forum for the exchange of information concerning the current state-of-the-art and potential of MAT and MT.

3. In the Intelligence Annex to the FY 78-82 Planning and Programming Guidance Memorandum (PPGM), 13 Feb 1976, the Deputy Secretary of Defense tasked the Air Force to: "chair studies which will determine current state of development of machine translation and its usefulness in the intelligence process." The PPGM further tasked the Army, Navy, NSA and DIA to participate in this effort and invited CIA to send representation. If this study determined that automated translation systems are efficient and economical, such findings together with proposed resource levels for a five year program should be furnished in a report to the Assistant Secretary of Defense (Intelligence).

4. In order to accomplish the PPGM tasking, on 19 Mar 1976 the Air Force convened a meeting of senior representatives from the Services and Agencies concerned to develop terms of reference for the study. A series of working group meetings were held in March and April to evaluate the state-of-the-art, refine Service/Agency requirements, and develop a program for MAT/MT. The DIA MT Survey and the data acquired through the FBIS Seminar provided valuable material for consideration by the MT Study Group. This report sets forth the findings and proposals of the Study Group.

#### CURRENT SITUATION:

1. FTD is the only DOD organization currently employing a large-scale MT system. The system is used successfully for translation of S&T materials. Output is provided in an unedited, partially edited, or fully edited version, depending on the requirements of the consumer. The system can provide indicative translations of journalistic/literary material but has never been optimized for such prose. FTD is also pursuing initiatives to improve the efficiency of the input and post-edit processes.

2. Other agencies employ human translation and rely principally on the Joint Publications Research Service (JPRS) and/or commercial vendors. The quality of some translation support is often so inferior that considerable additional editing is required before publication of the finished product.

a. Two major translation efforts are conducted by the Air Staff (Directorate of Threat Applications). They are: Monthly Soviet Press Translations and the Soviet Military Thought series, generally published in book form. These projects require translations to be in high quality, idiomatic English. The Soviet Military Thought series is an open-ended project, and each book averages 75,000 words in length. The current average monthly volume of Soviet press translations is 13,000 words. The Air Staff has identified requirements for translation of 13 additional Soviet journals and newspapers, totalling 14,000 pages annually.

b. Because of a present estimated human translation (HT) capability for the production per 12-month period of 17.5 million words from Russian, the Army OACSI is a long way from requiring MT (6.8 million words were the requirement produced from Russian by HT in FY 75). With an expansion of the currently small reserve-officer translation program, and the assignment of more projects of relatively less

difficulty to the Regular Army linguist units at Forts Bragg and Hood, an HT production. Approved For Release 2007/10/19 : CIA-RDP83-00714R000100410001-1 quite feasible by FY 78. Although 65-70 percent of the Russian material translated may be considered S&T (75-80 percent for all languages), such material usually comes in written context. With the frequent Intelligence Community and DOD-wide dissemination of OACSI translations, camera-ready copy entailing terminological accuracy, readability and graphics is ordinarily imperative. An ever growing roster of qualified commercial translation sources with subject specialties generally enables immediate assignment of, and good to excellent turnaround time for, translation projects.

c. The Translation Services Division of the Naval Intelligence Support Center translates a total of approximately 5 million words per year of carefully screened foreign literature. High quality is desired. The language breakdown is as follows: S&T, 3 million words; Naval, 1.5 million words; other 750,000 words. Russian accounts for 70 percent of the total volume, with Japanese, German, French and Italian accounting for most of the remaining 30 percent, in descending order. The Division regularly exploits 40 high-yield periodicals and newspapers and prepares abstracts and translation of tables of contents of about 220 books per year, which in turn generate requests for translations. Work is farmed out to individual translators whose product must meet Division quality standards. Individual consultants are brought in as needed to provide services in languages not represented in staff capabilities. The Translation Services Division operates a Reserve Translation Project which utilizes the linguistic skills of 70 Reserve officers and enlisted men in Russian, German, French, and Spanish. Applicants for the Program must pass a difficult test before they are accepted for participation. Current Navy MT-related initiatives are directed toward the development of specialized lexical aids. Navy requirements for socio-political literature is nearly completely satisfied by the FBIS/JPRS exploitation effort. Navy estimates that about 2 million words per year would be translated for intelligence exploitation if additional resources were available. That translating would represent "nice-to-have" material and would not have to meet Navy's quality standards.

d. DIA supports the MT efforts at FTD within the general area of the DOD Scientific and Technical Intelligence Information Support Program (STIISP). In addition, DIA is presently translating, or having translated for it, a total of approximately one million words per year. If additional translation capability were developed, DIA estimates that this requirement would increase to 1.9 million words per year.

e. FBIS currently produces 235,000 pages (100 million words) annually. Sixty languages are involved, with Russian accounting for 45 percent of the total workload. Approximately one-third of the total effort involves S&T material; the remaining two-thirds involves political, military, economic, biographic, and sociological material. FBIS maintains in-house a staff of linguists and draws on a roster of about [ ] translators under contract to JPRS. All translations must be of literary quality. FBIS maintains that adequate human resources are available to satisfy its requirements. MAT techniques can materially improve the quality and efficiency

of human translation and initiatives are being pursued in the development of lexical aids. ] Approved For Release 2007/10/19 : CIA-RDP83-00714R000100410001-1 : cost-effective programs are implemented.

f. NSA is extremely interested in the processing of natural-language material in both graphemic and phonemic form. Not all this material need be translated. However, what must be translated into English must eventually pass through several layers of quality control. The final output of the translation process, in addition to appearing with due timeliness, must maintain the semantics of the original material: it must omit nothing of significance, it must add nothing of significance and, to the greatest extent possible, it must minimize the distortion unavoidably resulting from conversion of semantics of the source language to semantics of the the target language. Currently, NSA is continuing its development of computerized lexical aids, dictionaries, and keyword search techniques. Recognizing the similarity of some of their needs in these areas of MAT, the NSA and CIA contingents to the MT Working Group have agreed to coordinate their MAT efforts.

g. The Intelligence Community Staff is concerned with the overall problem of linguistic expertise in the United States. This problem results from the relative lack of emphasis on foreign language training in the American academic environment and inadequate professional opportunities and rewards for linguists. The IC Staff supports initiatives in MAT to improve the competence and professional status of linguists and intelligence analysts and in MT to provide responsive translations of pertinent material for intelligence exploitation.

#### DISCUSSION:

1. Information provided by the DIA MT Survey indicates that there is a definite requirement for additional translation of Soviet material, principally in the journalistic/literary category of contemporary Soviet military doctrine, concepts and related subjects. It is not possible to determine how much of this additional requirement is duplicative, but the shortfall ranges from a minimum of 5 million word annually to a cumulative total of 12 million words. The actual requirement, therefore, is between these extremes and is, in any case, substantial.

2. The FTD experience is useful in determining the cost-effectiveness and utility of MT for high volume translation. At the request of the Assistant Secretary of the Air Force, Research and Development, the USAF Scientific Advisory Board (SAB) conducted a study of the FTD Translation System. The SAB reported (5 Jun 1975) that the output was highly acceptable to FTD analysts. The system was competitive in cost with human translation. The SAB noted that 75 percent of the MT cost was accounted for by post-editing and recomposition to provide material that is camera-ready for the printer. Automation of these processes, as planned by FTD, would lower the cost of a finished MT product below that of human translation. Subsequent system improvements have enabled FTD to provide analysts with a greater proportion of low cost unedited or partially edited MT output which satisfies user requirements.

a. Approved For Release 2007/10/19 : CIA-RDP83-00714R000100410001-1  
Russian material (each page averages 250 words), for the period 1970 through  
Mar 1976.

Unedited MT	20,991
Partially Edited MT	14,143
Finished MT	951
Manual (HT)	<u>5,013</u>
Total	41,098

b. The following figures represent approximate direct labor and materials cost per 1,000 words translated at FTD.

Unedited MT	\$ 8.63
Partially Edited MT	17.87
Finished MT	32.38
Manual Draft (HT)	27.28
Manual Finished (HT)	36.00

3. The Air Staff has conducted experiments using the FTD MT System for translation of purely journalistic/literary material. The output provided was indicative of content but would require excessive post-editing to obtain a literary English language product. The system's limitations in this area are accentuated because optimization efforts have never been directed toward a journalistic/literary capability. Upgrading any currently operational system is probably not the best approach to achieve an MT capability for literary quality output. Such systems are implemented on a conceptual base that is 20 years old. As stated by several participants and commentators at the FBIS MT Seminar, a fundamentally new approach is probably required to resolve the problems which have precluded current MT systems from providing a high quality output with minimal post-editing.

4. MT Working Group participants have expressed the requirements of their respective agencies for high quality literary translations. Admittedly, material of long term value intended for wide distribution should be published as a quality product. However, considerations involving the American-Soviet Copyright Agreement of May 1973 may well require the restriction of a large volume of material for internal government use only. In addition, there is a considerable amount of material of a transient nature which is required by intelligence analysts but need not be provided in a high quality or camera-ready form. A major constraint in using state-of-the-art MT systems for such timely indicative translations is the requirement for manual input of the material to be translated. It is currently more practical to have human translators scan material for content and value to intelligence analysis. Inasmuch as translation resources are limited within the Department of Defense and a volume of material with potential intelligence payoff goes untranslated, automation of the MT input process is a priority requirement. The development of an OCR system would bring the computer power of an MT system to bear on the problem in a cost-effective



Approved For Release 2007/10/19 : CIA-RDP83-00714R000100410001-1 : performance of the present L<sub>1</sub> system as well as for any third generation system which may evolve. Although the material on OCR technology presented at the FBIS MT Seminar was not encouraging, other opinions obtained from Seminar commentators indicated that such technology is approaching a stage where it can be successfully employed in an MT process.

5. A program to attain high volume, timely production of journalistic/literary Russian translations should have the overall aim of delivering the highest quality translation necessary and sufficient to satisfy the user's information requirements at the lowest cost. High quality translations have been the goal in the past without regard to whether such quality is in fact required in all cases for the user to perform his task. The user presumably is the expert in the discipline of the document and does contribute something to the interface between himself and the translation in comprehending the material. The validity of this concept is demonstrated by the fact that users at FTD and Oak Ridge National Laboratories use the raw output of their respective MT systems.

6. It is agreed that no MT system will totally replace the human translator. The goal in developing a third-generation MT system is to provide an output in idiomatic English that is faithful to the source input in content and meaning, with a minimum of human editing (5-10 percent). Such a system should provide the option for human intervention during the translation process as well as in a post-edit mode. The system should be modular and designed for ease of software maintenance. Finally, it should be as language-independent as possible to facilitate implementation of MT for languages other than Russian. Such a system would accommodate the requirement for timely translation of an increasing volume of pertinent material. By minimizing manpower-intensive input and editing functions, it would provide quality translations at a cost somewhere between that for unedited and partially edited MT at FTD.

7. There are numerous efforts underway which could lead to a third-generation MT system. Because immediate substantial payoff from investment in this technology is unlikely and the exact direction that the development effort should take is uncertain, a cautious and evolutionary approach is required. However, DOD involvement in such technological development is essential to insure that it is responsive to identified requirements. In this regard, near-term emphasis should be placed on MAT techniques, which offer more immediate practical benefits and which might provide a valuable contribution (e.g., through dictionary development) to any future MT system.

8. The MT Working Group participants agree that such initiatives in MAT methodologies should be pursued. Specifically, the emphasis should be placed on the continued development of dictionaries and lexical aids. Standard MAT software should be developed which would provide a common format for dictionary entries, provide on-line and batch processing capabilities for dictionary update and retrieval, and aids for editing and formatting of translations. Much of the dictionary development envisioned by Army and Navy will also contribute to improving the capability of the FTD system. The development of such MAT capabilities are well within

the current state-of-the-art and will contribute substantially toward increasing the pro. Approved For Release 2007/10/19 : CIA-RDP83-00714R000100410001-1 the FBIS MT Seminar was that MAT techniques could provide an increase in productivity by a factor of 2-1 or 3-1. It appears, however, that an increase of 60 percent is a more realistic assessment. In any event, the benefits are substantial.

9. Technological development and implementation have often been characterized by multiple independent efforts which result in duplicative capabilities and unnecessary costs. To avoid this situation in proposed MAT and MT development efforts, a formal coordinating structure should be established at the USIB level. In the interim, if so directed by OASD(I), the present Ad Hoc MT Working Group can perform this function for the DOD. In view of the difficulty in identifying the most lucrative approach to a follow-on MT capability, some mechanism for providing professional advice on MT development should be established. Considerable expertise is available at RADC. In addition, a carefully selected advisory body composed of experts from such relevant disciplines as linguistics, computational linguistics, computer science, psychology, human factors engineering, and artificial intelligence would be helpful. The Working Group recognizes the biases that exist in all of these disciplines and emphasizes the advisory nature of such a body of experts.

#### ✓ CONCLUSIONS:

- In the intelligence process, translations are principally useful insofar as the material translated contributes to the analysis of foreign capabilities and intentions. In this regard, considerations of comprehensiveness and timeliness must be weighed against requirements for quality that will insure the proper transfer of concept from one language into another.
- MT has proven cost-effective and responsive to some S&T user requirements.
- The current state-of-the-art of MT will not support quality production of journalistic/literary material without excessive post-editing. Lack of automated input technology precludes its effective use for timely indicative translations.
- Immediate benefits can be obtained from implementation of MAT methodologies and may also contribute to development of an advanced MT capability.
- A long-term, cautious and evolutionary development effort might provide a cost-effective system capable of providing timely, quality translations of needed materials, some of which is currently untranslated and probably unexploited.

#### ✓ RECOMMENDATIONS:

- That ASD(I) provide the following funding for implementation of near-term MAT and long-term OCR and journalistic/literary MT capabilities (\$ in thousands):

<u>FY77</u>	<u>FY78</u>	<u>FY79</u>	<u>FY80</u>	<u>FY81</u>	<u>FY82</u>

Approved For Release 2007/10/19 : CIA-RDP83-00714R000100410001-1 /MT  
- That ASD(1) establish a formal coordinating group with OS for MAT/MT  
and designate an Executive Agent for MAT/MT implementation.

- That a similar structure be established by the USIB to address and coordinate overall community translation requirements, including both the improvement of translator professionalism and the implementation of automated aids for translation.

- That RADC be tasked to further refine overall translation requirements, assist in development of Service/Agency Statements of Work for MAT/MT support, and identify the allocation of funding (by appropriation) needed for MAT/MT development and implementation.

Submitted by the Ad Hoc Machine Translation Working Group.

Air Force

Col W. P. Olsen	AFIS/IND
Col N. P. Vaslef	AF/INA
Maj R. E. Baldauf	AF/INY
Maj L. M. Hansen	FTD/NIT

Army

Mr. G. C. Cooney	OACSI
------------------	-------

Navy

Mr. T. P. Koines	NISC
Mr. C. R. Moctezuma	NISC

DIA

STAT  DIA/DT3

NSA

STAT  NSA/R51

✓ FBIS

STAT  FBIS/EPS

✓ I. C. Staff

STAT

APPROVED:

  
WILLIAM P. OLSEN, Colonel, USAF  
Chairman